



<400> 1

gcctggctcc ctctcgctga gacacacata cactcacaca tacacaaccc ggcaggctcg 60

tctgaacttg aagacacccc acattccaag atgcccgagg ttcttgggaa tgcctggggt 120

tcttcgatcc ggaaaatcct accggcatcc tcctagggag ggattattat tattatTTTT 180

ctttaatctg gaagagaaga gaacaagttg tgcttttccc ccctttctct tgctaaacgc 240

catggatata actgaataag cggctcaggg ctttcccgc gtggacgtcc gaggccacca 300

tctgcctgca ttgcgccgag ccgccggagg gtttagctcg agtctgtctc gggcggggaa 360

ggatgcgtgg ccgagccggg gagcccgggc gccccgcgga gccggcctcg gtgccacca 420

gccgggggta gatgctgcct cgcccaggcg ctgagtgacc agacc atg gag acc ctg 477

Met Glu Thr Leu

1

ctt ggt ggc ctg cta gcg ttt ggc atg gcg ttt gcc gtg gtc gac gcc 525

Leu Gly Gly Leu Leu Ala Phe Gly Met Ala Phe Ala Val Val Asp Ala

5 10 15 20

tgc ccc aag tac tgt gtc tgc cag aat ctg tct gag tca ctg ggg acc 573

Cys Pro Lys Tyr Cys Val Cys Gln Asn Leu Ser Glu Ser Leu Gly Thr

25 30 35

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted February 19, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ctg tgc ccc tcc aag ggg ctg ctc ttt gta ccc cct gat att gac cgg	621
Leu Cys Pro Ser Lys Gly Leu Leu Phe Val Pro Pro Asp Ile Asp Arg	
40 45 50	
cgg aca gtg gag ctg cgc ctg ggc ggc aac ttc atc atc cac atc agc	669
Arg Thr Val Glu Leu Arg Leu Gly Gly Asn Phe Ile Ile His Ile Ser	
55 60 65	
cgc cag gac ttt gcc aac atg acg ggg ctg gtg gac ctg acc ctg tcc	717
Arg Gln Asp Phe Ala Asn Met Thr Gly Leu Val Asp Leu Thr Leu Ser	
70 75 80	
agg aac acc atc agc cac atc cag ccc ttt tcc ttt ctg gac ctc gag	765
Arg Asn Thr Ile Ser His Ile Gln Pro Phe Ser Phe Leu Asp Leu Glu	
85 90 95 100	
agc ctc cgc tcc ctg cat ctt gac agc aat cgg ctg cca agc ctt ggg	813
Ser Leu Arg Ser Leu His Leu Asp Ser Asn Arg Leu Pro Ser Leu Gly	
105 110 115	
gag gac acc ctc cgg ggc ctg gtc aac ctg cag cac ctt atc gtg aac	861
Glu Asp Thr Leu Arg Gly Leu Val Asn Leu Gln His Leu Ile Val Asn	
120 125 130	
aac aac cag ctg ggc ggc atc gca gat gag gct ttt gag gac ttc ctg	909

Asn Asn Gln Leu Gly Gly Ile Ala Asp Glu Ala Phe Glu Asp Phe Leu	
135	140
ctg aca ttg gag gat ctg gac ctc tcc tac aac aac ctc cat ggc ctg	957
Leu Thr Leu Glu Asp Leu Asp Leu Ser Tyr Asn Asn Leu His Gly Leu	
150	155
160	
ccg tgg gac tcc gtg cga cgc atg gtc aac ctc cac cag ctg agc ctg	1005
Pro Trp Asp Ser Val Arg Arg Met Val Asn Leu His Gln Leu Ser Leu	
165	170
175	180
gac cac aac ctg ctg gat cac atc gcc gag ggc acc ttt gca gac ctg	1053
Asp His Asn Leu Leu Asp His Ile Ala Glu Gly Thr Phe Ala Asp Leu	
185	190
195	
cag aaa ctg gcc cgc ctg gat ctc acc tcc aat cgg ctg cag aag ctg	1101
Gln Lys Leu Ala Arg Leu Asp Leu Thr Ser Asn Arg Leu Gln Lys Leu	
200	205
210	
ccc cct gat ccc atc ttt gcc cgc tcc cag gct tcg gct ttg aca gcc	1149
Pro Pro Asp Pro Ile Phe Ala Arg Ser Gln Ala Ser Ala Leu Thr Ala	
215	220
225	
aca ccc ttt gcc cca ccc ttg tcc ttt agt ttt ggg ggt aac cca ctt	1197
Thr Pro Phe Ala Pro Pro Leu Ser Phe Ser Phe Gly Gly Asn Pro Leu	
230	235
240	

1000  
 1001  
 1002  
 1003  
 1004  
 1005  
 1006  
 1007  
 1008  
 1009  
 1010  
 1011  
 1012  
 1013  
 1014  
 1015  
 1016  
 1017  
 1018  
 1019  
 1020  
 1021  
 1022  
 1023  
 1024  
 1025  
 1026  
 1027  
 1028  
 1029  
 1030  
 1031  
 1032  
 1033  
 1034  
 1035  
 1036  
 1037  
 1038  
 1039  
 1040  
 1041  
 1042  
 1043  
 1044  
 1045  
 1046  
 1047  
 1048  
 1049  
 1050  
 1051  
 1052  
 1053  
 1054  
 1055  
 1056  
 1057  
 1058  
 1059  
 1060  
 1061  
 1062  
 1063  
 1064  
 1065  
 1066  
 1067  
 1068  
 1069  
 1070  
 1071  
 1072  
 1073  
 1074  
 1075  
 1076  
 1077  
 1078  
 1079  
 1080  
 1081  
 1082  
 1083  
 1084  
 1085  
 1086  
 1087  
 1088  
 1089  
 1090  
 1091  
 1092  
 1093  
 1094  
 1095  
 1096  
 1097  
 1098  
 1099  
 1100



Asn Gly Thr Leu Asp Ile Phe Ile Thr Thr Ser Gln Asp Ser Gly Ala  
 345 350 355

ttc acc tgc att gct gcc aat gct gcc gga gag gcc acg gcc atg gtg 1581  
 Phe Thr Cys Ile Ala Ala Asn Ala Ala Gly Glu Ala Thr Ala Met Val  
 360 365 370

gag gtc tcc atc gtc cag ctg cca cac ctc agc aac agc acc agc cgc 1629  
 Glu Val Ser Ile Val Gln Leu Pro His Leu Ser Asn Ser Thr Ser Arg  
 375 380 385

act gca ccc ccc aag tcc cgc ctc tca gac atc act ggc tcc agc aag 1677  
 Thr Ala Pro Pro Lys Ser Arg Leu Ser Asp Ile Thr Gly Ser Ser Lys  
 390 395 400

acc agc cgg gga ggt gga ggc agt ggg ggc gga gag cct ccc aaa agc 1725  
 Thr Ser Arg Gly Gly Gly Gly Ser Gly Gly Gly Glu Pro Pro Lys Ser  
 405 410 415 420

ccc ccg gaa cgg gct gtg ctt gtg tct gaa gtg acc acc acc tcg gcc 1773  
 Pro Pro Glu Arg Ala Val Leu Val Ser Glu Val Thr Thr Thr Ser Ala  
 425 430 435

ctg gtc aag tgg tct gtc agc aag tca gca ccc cgg gtg aag atg tac 1821  
 Leu Val Lys Trp Ser Val Ser Lys Ser Ala Pro Arg Val Lys Met Tyr  
 440 445 450

1581 1629 1677 1725 1773 1821

cag ctg cag tac aac tgc tct gac gat gag gta ctg att tac agg atg	1869
Gln Leu Gln Tyr Asn Cys Ser Asp Asp Glu Val Leu Ile Tyr Arg Met	
455 460 465	
atc cca gcc tcc aac aag gcc ttc gtg gtc aac aac ctg gtg tca ggg	1917
Ile Pro Ala Ser Asn Lys Ala Phe Val Val Asn Asn Leu Val Ser Gly	
470 475 480	
act ggc tac gac ttg tgt gtg ctg gcc atg tgg gat gac aca gcc acg	1965
Thr Gly Tyr Asp Leu Cys Val Leu Ala Met Trp Asp Asp Thr Ala Thr	
485 490 495 500	
aca ctc acg gcc acc aac atc gtg ggc tgc gcc cag ttc ttc acc aag	2013
Thr Leu Thr Ala Thr Asn Ile Val Gly Cys Ala Gln Phe Phe Thr Lys	
505 510 515	
gct gac tac ccg cag tgc cag tcc atg cac agc cag att ctg ggc ggc	2061
Ala Asp Tyr Pro Gln Cys Gln Ser Met His Ser Gln Ile Leu Gly Gly	
520 525 530	
acc atg atc ctg gtc atc ggg ggc atc atc gtg gcc acg ctg ctg gtc	2109
Thr Met Ile Leu Val Ile Gly Gly Ile Ile Val Ala Thr Leu Leu Val	
535 540 545	
ttc atc gtc atc ctc atg gtg cgc tac aag gtc tgc aac cac gag gcc	2157

Phe Ile Val Ile Leu Met Val Arg Tyr Lys Val Cys Asn His Glu Ala  
 550 555 560

ccc agc aag atg gca gcg gcc gtg agc aat gtg tac tcg cag acc aac 2205  
 Pro Ser Lys Met Ala Ala Ala Val Ser Asn Val Tyr Ser Gln Thr Asn  
 565 570 575 580

ggc gcc cag cca ccg cct cca agc agc gca cca gcc ggg gcc ccg ccg 2253  
 Gly Ala Gln Pro Pro Pro Pro Ser Ser Ala Pro Ala Gly Ala Pro Pro  
 585 590 595

cag ggc ccg ccg aag gtg gtg gtg cgc aac gag ctc ctg gac ttc acc 2301  
 Gln Gly Pro Pro Lys Val Val Val Arg Asn Glu Leu Leu Asp Phe Thr  
 600 605 610

gcc agc ctg gcc cgc gcc agt gac tcc tct tcc tcc agc tcc ctg ggc 2349  
 Ala Ser Leu Ala Arg Ala Ser Asp Ser Ser Ser Ser Ser Ser Leu Gly  
 615 620 625

agt ggg gag gct gcg ggg ctg gga cgg gcc ccc tgg agg atc cca ccc 2397  
 Ser Gly Glu Ala Ala Gly Leu Gly Arg Ala Pro Trp Arg Ile Pro Pro  
 630 635 640

tcc gcc ccg cgc ccc aag ccc agc ctt gac cgc ctg atg ggg gcc ttc 2445  
 Ser Ala Pro Arg Pro Lys Pro Ser Leu Asp Arg Leu Met Gly Ala Phe  
 645 650 655 660

2025 RELEASE UNDER E.O. 14176



gcc tcc ctg gac ctc aag agt cag aga aag gag gag ctg ctg gac tcc	2493
Ala Ser Leu Asp Leu Lys Ser Gln Arg Lys Glu Glu Leu Leu Asp Ser	
665 670 675	
agg act cca gcc ggg aga ggg gct ggg acg tcg gcc cgg ggc cac cac	2541
Arg Thr Pro Ala Gly Arg Gly Ala Gly Thr Ser Ala Arg Gly His His	
680 685 690	
tcg gac cga gag cca ctg ctg ggg ccc cct gcg gcc cgg gcc agg agc	2589
Ser Asp Arg Glu Pro Leu Leu Gly Pro Pro Ala Ala Arg Ala Arg Ser	
695 700 705	
ctg ctc ccc ttg ccg ttg gag ggc aag gcc aaa cgc agc cac tcc ttc	2637
Leu Leu Pro Leu Pro Leu Glu Gly Lys Ala Lys Arg Ser His Ser Phe	
710 715 720	
gac atg ggg gac ttt gct gct gcg gcg gcg gga ggg gtc gtg ccg ggc	2685
Asp Met Gly Asp Phe Ala Ala Ala Ala Ala Gly Gly Val Val Pro Gly	
725 730 735 740	
ggc tac agt cct cct cgg aag gtc tcg aac atc tgg acg aag cgc agc	2733
Gly Tyr Ser Pro Pro Arg Lys Val Ser Asn Ile Trp Thr Lys Arg Ser	
745 750 755	
ctc tct gtc aac ggc atg ctc ttg ccc ttt gag gag agt gac ctg gtg	2781

Leu Ser Val Asn Gly Met Leu Leu Pro Phe Glu Glu Ser Asp Leu Val

760

765

770

ggg gcc cgg ggg act ttt ggc agc tcc gaa tgg gtg atg gag agc acg 2829

Gly Ala Arg Gly Thr Phe Gly Ser Ser Glu Trp Val Met Glu Ser Thr

775

780

785

gtc taggtggggg tgggcatgct ccctttcctg tgcgcagggt gggagaagg 2882

Val

gaaagaatct cactggcaag tgtttgtgga gtttccatgg tgatgtttac atccaggac 2942

agtttcgtct ccctgtcaat ggctctgtgt cccccctac cccgcaacac ccacatcacc 3002

tccccaccac cggccgggg tgtgctcagg gaatgtggac tcgctcaaat gccggactga 3062

gccctgagtg tttggaaagg cgagactccg cctttctaata cacaaatgta gcctacaagc 3122

aagcggcttt ggattgctta tg 3144

<210> 2

<211> 789

<212> PRT

<213> Homo sapiens

2829 2942 3002 3062 3122 3144

&lt;400&gt; 2

Met Glu Thr Leu

1

Leu Gly Gly Leu Leu Ala Phe Gly Met Ala Phe Ala Val Val Asp Ala

5 10 15 20

Cys Pro Lys Tyr Cys Val Cys Gln Asn Leu Ser Glu Ser Leu Gly Thr

25 30 35

Leu Cys Pro Ser Lys Gly Leu Leu Phe Val Pro Pro Asp Ile Asp Arg

40 45 50

Arg Thr Val Glu Leu Arg Leu Gly Gly Asn Phe Ile Ile His Ile Ser

55 60 65

Arg Gln Asp Phe Ala Asn Met Thr Gly Leu Val Asp Leu Thr Leu Ser

70 75 80

Arg Asn Thr Ile Ser His Ile Gln Pro Phe Ser Phe Leu Asp Leu Glu

85 90 95 100

Ser Leu Arg Ser Leu His Leu Asp Ser Asn Arg Leu Pro Ser Leu Gly

105 110 115

LEU-GLY-GLY-LEU-LEU-ALA-PHE-GLY-MET-ALA-PHE-ALA-VAL-VAL-ASP-ALA

Glu Asp Thr Leu Arg Gly Leu Val Asn Leu Gln His Leu Ile Val Asn  
 120 125 130

Asn Asn Gln Leu Gly Gly Ile Ala Asp Glu Ala Phe Glu Asp Phe Leu  
 135 140 145

Leu Thr Leu Glu Asp Leu Asp Leu Ser Tyr Asn Asn Leu His Gly Leu  
 150 155 160

Pro Trp Asp Ser Val Arg Arg Met Val Asn Leu His Gln Leu Ser Leu  
 165 170 175 180

Asp His Asn Leu Leu Asp His Ile Ala Glu Gly Thr Phe Ala Asp Leu  
 185 190 195

Gln Lys Leu Ala Arg Leu Asp Leu Thr Ser Asn Arg Leu Gln Lys Leu  
 200 205 210

Pro Pro Asp Pro Ile Phe Ala Arg Ser Gln Ala Ser Ala Leu Thr Ala  
 215 220 225

Thr Pro Phe Ala Pro Pro Leu Ser Phe Ser Phe Gly Gly Asn Pro Leu  
 230 235 240

His Cys Asn Cys Glu Leu Leu Trp Leu Arg Arg Leu Glu Arg Asp Asp  
 245 250 255 260

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

Asp Leu Glu Thr Cys Gly Ser Pro Gly Gly Leu Lys Gly Arg Tyr Phe

265

270

275

Trp His Val Arg Glu Glu Glu Phe Val Cys Glu Pro Pro Leu Ile Thr

280

285

290

Gln His Thr His Lys Leu Leu Val Leu Glu Gly Gln Ala Ala Thr Leu

295

300

305

Lys Cys Lys Ala Ile Gly Asp Pro Ser Pro Leu Ile His Trp Val Ala

310

315

320

Pro Asp Asp Arg Leu Val Gly Asn Ser Ser Arg Thr Ala Val Tyr Asp

325

330

335

340

Asn Gly Thr Leu Asp Ile Phe Ile Thr Thr Ser Gln Asp Ser Gly Ala

345

350

355

Phe Thr Cys Ile Ala Ala Asn Ala Ala Gly Glu Ala Thr Ala Met Val

360

365

370

Glu Val Ser Ile Val Gln Leu Pro His Leu Ser Asn Ser Thr Ser Arg

375

380

385

Thr Ala Pro Pro Lys Ser Arg Leu Ser Asp Ile Thr Gly Ser Ser Lys

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

390 395 400

Thr Ser Arg Gly Gly Gly Gly Ser Gly Gly Gly Glu Pro Pro Lys Ser  
405 410 415 420

Pro Pro Glu Arg Ala Val Leu Val Ser Glu Val Thr Thr Thr Ser Ala  
425 430 435

Leu Val Lys Trp Ser Val Ser Lys Ser Ala Pro Arg Val Lys Met Tyr  
440 445 450

Gln Leu Gln Tyr Asn Cys Ser Asp Asp Glu Val Leu Ile Tyr Arg Met  
455 460 465

Ile Pro Ala Ser Asn Lys Ala Phe Val Val Asn Asn Leu Val Ser Gly  
470 475 480

Thr Gly Tyr Asp Leu Cys Val Leu Ala Met Trp Asp Asp Thr Ala Thr  
485 490 495 500

Thr Leu Thr Ala Thr Asn Ile Val Gly Cys Ala Gln Phe Phe Thr Lys  
505 510 515

Ala Asp Tyr Pro Gln Cys Gln Ser Met His Ser Gln Ile Leu Gly Gly  
520 525 530

COLETTA

Thr Met Ile Leu Val Ile Gly Gly Ile Ile Val Ala Thr Leu Leu Val  
 535 540 545

Phe Ile Val Ile Leu Met Val Arg Tyr Lys Val Cys Asn His Glu Ala  
 550 555 560

Pro Ser Lys Met Ala Ala Ala Val Ser Asn Val Tyr Ser Gln Thr Asn  
 565 570 575 580

Gly Ala Gln Pro Pro Pro Pro Ser Ser Ala Pro Ala Gly Ala Pro Pro  
 585 590 595

Gln Gly Pro Pro Lys Val Val Val Arg Asn Glu Leu Leu Asp Phe Thr  
 600 605 610

Ala Ser Leu Ala Arg Ala Ser Asp Ser Ser Ser Ser Ser Ser Leu Gly  
 615 620 625

Ser Gly Glu Ala Ala Gly Leu Gly Arg Ala Pro Trp Arg Ile Pro Pro  
 630 635 640

Ser Ala Pro Arg Pro Lys Pro Ser Leu Asp Arg Leu Met Gly Ala Phe  
 645 650 655 660

Ala Ser Leu Asp Leu Lys Ser Gln Arg Lys Glu Glu Leu Leu Asp Ser  
 665 670 675

Arg Thr Pro Ala Gly Arg Gly Ala Gly Thr Ser Ala Arg Gly His His

680

685

690

Ser Asp Arg Glu Pro Leu Leu Gly Pro Pro Ala Ala Arg Ala Arg Ser

695

700

705

Leu Leu Pro Leu Pro Leu Glu Gly Lys Ala Lys Arg Ser His Ser Phe

710

715

720

Asp Met Gly Asp Phe Ala Ala Ala Ala Ala Gly Gly Val Val Pro Gly

725

730

735

740

Gly Tyr Ser Pro Pro Arg Lys Val Ser Asn Ile Trp Thr Lys Arg Ser

745

750

755

Leu Ser Val Asn Gly Met Leu Leu Pro Phe Glu Glu Ser Asp Leu Val

760

765

770

Gly Ala Arg Gly Thr Phe Gly Ser Ser Glu Trp Val Met Glu Ser Thr

775

780

785

Val



<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: an  
artificially synthesized primer sequence.

<400> 3

gactagttct agatcgcgag cggccgcct ttttttttt tttt

44

<210> 4

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: an  
artificially synthesized adapter sequence.

<400> 4

tcgacccacg cgtccg

16

<210> 5

<211> 12

<212> DNA

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: an  
artificially synthesized adapter sequence.

<400> 5

cggacgcgtg gg

12

<210> 6

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: an  
artificially synthesized primer sequence.

<400> 6

cagggtggga gaaggggaaa gaatc

25

<210> 7

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: an  
artificially synthesized primer sequence.

<400> 7

gaggccattg acagggagac gaaac

25

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100